

## Testen von Regressionskoeffizienten bei multipler Regression (ausführlichere Erläuterungen und Zahlenbeispiele)<sup>1</sup>

### 1. Übersicht

Gegeben ist die geschätzte Regressionsgleichung  $\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_K x_K$

(K ist hier die Anzahl der Regressoren, nicht die Anzahl der zu schätzenden Regressionskoeffizienten, die wegen  $\alpha$  bei dieser Notation K + 1 ist; darauf ist beim Vergleich der Formeln in diesem Papier mit entsprechenden Formeln an anderer Stelle, wo die Notation evtl. anders gewählt wurde, zu achten).

Die geschätzten Varianzen  $\hat{\sigma}_{\hat{\beta}_i}^2$  der Regressionskoeffizienten und damit auch die Standardabweichungen<sup>2</sup> ( $\hat{\sigma}_{\hat{\alpha}}, \hat{\sigma}_{\hat{\beta}_1}, \hat{\sigma}_{\hat{\beta}_2}, \dots$ ) werden benötigt, um Konfidenzintervalle zu berechnen und Hypothesentests durchzuführen. Im Falle von K = 2 Regressoren gilt<sup>1)</sup> (bei t-Verteilung mit T-3 Freiheitsgraden denn 3 = K + 1)

$$\sigma_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}^2}{S_{11}(1-R_{1,2}^2)} \text{ (und } \hat{\sigma}_{\hat{\beta}_2}^2 \text{ analog) mit } \hat{\sigma}^2 = \frac{S_{\hat{u}\hat{u}}}{T-3} \text{ (K = 2, } S_{\hat{u}\hat{u}} = \sum \hat{u}_t^2 \text{) sowie}$$

$$\hat{\sigma}_{\hat{\alpha}}^2 = \frac{\hat{\sigma}^2}{T} + (\bar{x}_1)^2 \hat{\sigma}_{\hat{\beta}_1}^2 + (\bar{x}_2)^2 \hat{\sigma}_{\hat{\beta}_2}^2 + 2\bar{x}_1 \bar{x}_2 \hat{\sigma}_{\hat{\beta}_1 \hat{\beta}_2}$$

Die folgenden Formeln für  $\hat{\sigma}_{\hat{\alpha}}^2$  und  $\hat{\sigma}_{\hat{\beta}}^2$  im Falle der einfachen linearen Regression (K = 1) sind leicht als spezielle Fälle hieraus herleitbar.

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{S_{xx}} \text{ mit } \hat{\sigma}^2 = \hat{\sigma}_u^2 = \frac{S_{\hat{u}\hat{u}}}{T-2} \text{ sowie } \hat{\sigma}_{\hat{\alpha}}^2 = \hat{\sigma}^2 \left( \frac{1}{T} + \frac{\bar{x}^2}{S_{xx}} \right) = \left( \frac{1}{T} \sum x_t^2 \right) \hat{\sigma}_{\hat{\beta}}^2$$

Bei der multiplen Regression sind jetzt folgende Tests möglich

1. Signifikanztests (**t-Tests**) über **einzelne Regressionskoeffizienten** bzw. **eine** Linearkombination von Regressionskoeffizienten.

a) Die Nullhypothese könnte dann lauten

$$H_0 : \beta_3 = 0 \text{ oder } \beta_2 = 0,5 \text{ oder } \beta_1 + \beta_2 = 0, \text{ so dass } \beta_1 = \beta_2$$

b) Die Nullhypothese ist *eine* Gleichung, die mehrere Regressionskoeffizienten umfasst. Die Hypothese  $\beta_1 + \beta_2 = 0$  wäre ein Beispiel für *eine* Linearkombination, die allgemein formuliert  $\mathbf{k}'\boldsymbol{\beta} = c$  in diesem Beispiel wären die Vektoren  $\mathbf{k}' = [1 \quad -1]$   $\boldsymbol{\beta}' = [\beta_1 \quad \beta_2]$  und der Skalar  $c = 0$ ).

Und eine Art von Signifikanztests, die es bei einfacher Regression noch nicht gibt:

2. Signifikanztests (**F-Tests**) über **einige bzw. alle Regressionskoeffizienten**

a) alle: Signifikanz des Gesamtzusammenhangs (K Regressoren)<sup>3</sup>  $H_0 : \beta_1 = \dots = \beta_K = 0$

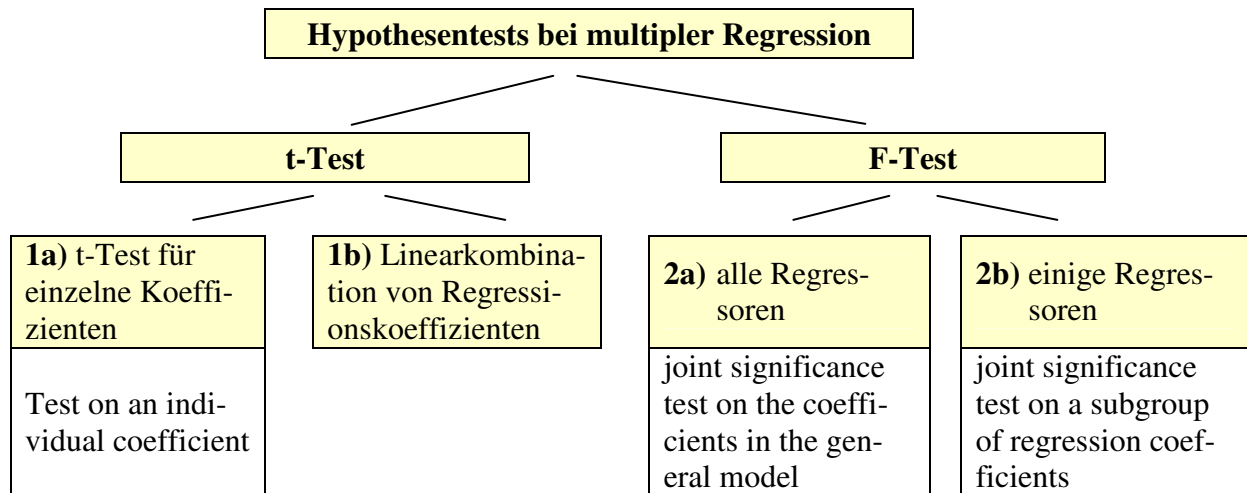
<sup>1</sup> Vgl. Download "Übersicht zur Schätz- und Testtheorie in der Regressionsanalyse". Im Grunde sind dort alle Formeln und Vorgehensweisen, die hier erläutert werden, auch bereits genannt und knapp dargestellt. Der hier vorliegende Download ist dazu nur eine weitere Erläuterung und Veranschaulichung mit Zahlenbeispielen.

<sup>2</sup> standard deviations, oder standard errors

<sup>3</sup> Das Absolutglied  $\alpha$  wird hier nicht betrachtet.

- b) einige (mehrere): sind die (zusätzlichen) Regressoren  $x_3$  und  $x_4$  signifikant? (über  $x_1$  und  $x_2$  wird keine Hypothese gebildet)  $H_0: \beta_3 = \beta_4 = 0$  ( $\beta_1$  und  $\beta_2$  können dann *irgendwelche* Werte haben) Bei Geltung von  $H_0$  hat das Modell jetzt  $L = 2$  (Anzahl der Gleichheitszeichen zählen!) Restriktionen (oder "Linearkombinationen").

*Man erhält somit das folgende Schema*



## 2. Der t-Test (und das Konfidenzintervall) für einen einzelnen Regressionskoeffizienten (Fall 1a)

Bei der Hypothese  $H_0: \beta_i = c$  ist die Prüfgröße  $t = \frac{\hat{\beta}_i - c}{\hat{\sigma}_{\beta_i}} \sim t_{T-K-1}$  (äquivalent  $t^2 \sim F_{1, T-K-1}$ )<sup>4</sup>

**Zahlenbeispiel** (Affenaufgabe)  $T = 5$

Geschätzte Gleichung  $\hat{y}_t = -28 + 2,25 x_t$ ,  $S_{\hat{u}\hat{u}} = 349$ ,  $\hat{\sigma}^2 = S_{\hat{u}\hat{u}} / (T - 2) = 349 / 3 = 116,33$

$$\hat{V} = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 404,842 & -13,24 \\ -13,24 & 0,460 \end{bmatrix}, r = 0,8864, r^2 = 0,7857, S_{yy} = 1628,8$$

$H_0: \beta = 0 \Rightarrow$  Prüfgröße  $\Rightarrow t = \frac{\hat{\beta} - 0}{\hat{\sigma}_{\beta}} = \frac{2,25}{\sqrt{0,46}} = 3,3173$ , der Tabellenwert der t-Verteilung bei

$T - 2 = 3$  Freiheitsgraden und einem Signifikanzniveau von 5 % zweiseitig ist 2,3534 also ist  $\hat{\beta} = 2,25$  gegen  $\beta = 0$  gesichert, da  $3,3173 > 2,3534$ .

Die Grenzen des 95 % Konfidenzintervalls sind dann entsprechend bei  $\beta$

$$2,25 \pm 2,3534 \cdot \sqrt{0,46} \quad \text{und bei } \alpha \text{ sind die Grenzen } -28 \pm 2,3534 \cdot \sqrt{404,842}.$$

<sup>4</sup> Der Test läuft also auf das Gleiche hinaus wie ein F-Test mit der Prüfgröße  $t^2$ , die dann F-verteilt ist mit einem und  $T - K - 1$  Freiheitsgraden (vgl. Abschn. 4b).

### 3. Testen einer (L = 1) Linearkombination (linearen Restriktion) mit dem t-Test (Fall 1b)

$H_0 : r_1 \beta_1 + r_2 \beta_2 = q$  (Beispiel aus dem Buch von L.v.Auer, S. 194)<sup>5</sup>.

Die geschätzte Standardabweichung von dieser Linearkombination<sup>6</sup>  $r' \hat{\beta} = \begin{bmatrix} r_1 & r_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = q$  ist

dann  $se = \left[ r_1^2 \frac{\hat{\sigma}^2}{S_{11}(1-R_{.21}^2)} + r_2^2 \frac{\hat{\sigma}^2}{S_{22}(1-R_{1.2}^2)} - 2r_1r_2 \frac{\hat{\sigma}^2 R_{1.2}^2}{S_{12}(1-R_{1.2}^2)} \right]^{1/2}$  mit  $\hat{\sigma}^2 = \frac{S_{\hat{u}\hat{u}}}{T-3}$  und die Prüf-

größe lautet dann  $t = \frac{r' \hat{\beta} - q}{se} \sim t_{T-K-1}$ .

Auch dieser Test ist wieder äquivalent eines Tests der Prüfgröße  $t^2$  mit der F-Verteilung bei  $L=1$  und  $T-K-1$  Freiheitsgraden gem. Abschn. 5.

Auch das isolierte Testen eines Regressors etwa  $\beta_2$  (Abschn. 2) ist ein Spezialfall dieses Abschnitts. Dann ist  $r_1 = 0$  und  $r_2 = 1$ . Die (geschätzte) Standardabweichung  $se = \hat{\sigma}_{\hat{\beta}_2}$  ist dann einfach  $[\hat{\sigma}^2 / S_{22}(1-R_{1.2}^2)]^{1/2}$ .

**Zahlenbeispiele** für diesen Test sind in der Literatur nicht leicht zu finden, weil der Zahlenwert für die Kovariation  $S_{12} = \sum (x_{1t} - \bar{x}_1)(x_{2t} - \bar{x}_2)$  meist nicht mitgeteilt wird. Im Folgenden ein Beispiel aus Murphy, Introductory Econometrics, p. 219

$$\hat{y} = 2,5824 - 1,00494 x_1 + 1,52021 x_2 \quad \text{mit } T = 24, K = 2.$$

Für die Summen der (quadrierten) Abweichungen erhält man  $S_{11} = 4703,983$

$S_{22} = 1113,062$  und  $S_{12} = 2163,541$ . Die Hypothese könnte hier lauten  $\beta_2 = -\beta_1$  oder

$H_0 : \beta_1 + \beta_2 = 0$ . Für die Standardabweichung der Linearkombination erhält man bei

$$\hat{\sigma}^2 = 17,838 \text{ und } R_{1.2}^2 = r_{1.2}^2 = 0,2880$$

$$se = \left[ \frac{17,838}{0,712} \left( \frac{1}{4704} + \frac{1}{1113} - \frac{2 \cdot 0,288}{2163,5} \right) \right]^{1/2} = 0,1455 \text{ und die berechnete Prüfgröße ist dann}$$

$$t = \frac{(-1,00494 + 1,52021) - 0}{0,1455} = 3,5419. \text{ Sie ist mit dem Tabellenwert der t-Verteilung zu}$$

vergleichen.

Der Tabellenwert der t-Verteilungstabelle ist bei  $T-K-1 = 24 - 2 - 1 = 21$  Freiheitsgraden und 5% einseitig 2,08, so dass  $H_0$  abzulehnen ist. Auch bei einer größeren Stichprobe nicht signifikant, weil der entsprechende Wert der Standardnormalverteilung 1,6449 beträgt. Mit der Hypothese  $H_0 : \beta_1 + \beta_2 = 0,5$  erhielte man jedoch einen viel kleineren Zähler in der Prüfgröße  $t$ , nämlich 0,01527 und damit  $t = 0,10496$ , was  $< 2,08$  ist, also *diese*  $H_0$  könnte man nicht verwerfen.

<sup>5</sup>  $q$  ist hier eine Konstante wie  $c$ .

<sup>6</sup> Eine spezielle Linearkombination dieser Art liegt auch den bekannten Formeln für das Prognoseintervall bei linearer einfacher Regression zugrunde mit  $k_1=1$  und  $k_2= x_0$ , sowie dem Koeffizientenvektor  $[\alpha \ \beta]$ , also mit dem Modell  $\alpha + \beta x_0$ .

#### 4. Der F-Test für Signifikanz aller Regressoren (Fall 2a)

##### a) Zweifache Regression

Demonstriert wieder mit dem Beispiel aus Murphy, Introductory Econometrics, p. 219

$\hat{y} = 2,5824 - 1,00494 x_1 + 1,52021 x_2$  mit  $T = 24$ ,  $K = 2$  (Anzahl der Regressoren), die geschätzten Standardabweichungen sind  $\hat{\sigma}_{\beta_1} = 0,18917$  und  $\hat{\sigma}_{\beta_2} = 0,38888$ . Die t-Werte  $\hat{\beta} / \hat{\sigma}_{\beta}$  sind 5,3124 bei  $\beta_1$  und 3,9092 bei  $\beta_2$  (entsprechend nimmt  $F = t^2$  die Werte 28,22 bzw. 15,28 an). Bei diesen Werten für t sind die Regressionskoeffizienten für sich genommen jeweils gegen  $\beta_1 = 0$  und  $\beta_2 = 0$  gesichert.

Sind sie auch zusammen genommen von Null verschieden? Kann also auch  $H_0 : \beta_1 = \beta_2 = 0$  (mit  $L = 2$  Restriktionen) verworfen werden?

Dazu ist die folgende Tabelle zu berechnen (MSE = mean square error = Variation / Anzahl der Freiheitsgrade)<sup>7</sup>

Variation	Freiheitsgrade	MSE
<b>erklärt</b> $S_{\hat{y}\hat{y}} = 712,296$ (Regression)	$K = 2$	$\frac{712,296}{2} = 356,148$
<b>residual</b> $S_{\hat{u}\hat{u}} = 374,595$ (nicht erklärt)	$T - K - 1 = 21$	$\frac{374,595}{21} = 17,838$
<b>total</b> $S_{yy} = 1086,891$	$T - 1 = 23$	

Die F-verteilte Prüfgröße ist dann

$$F = \frac{\text{erklärte Varianz}}{\text{Residual Varianz}} = \frac{S_{\hat{y}\hat{y}} / K}{S_{\hat{u}\hat{u}} / (T - K - 1)} \sim F_{K, T-K-1}$$

Im Beispiel ist  $F = 356,148 / 17,838 = 19,966$  und der Wert in der F-Tabelle bei 5 % Signifikanzniveau und 2 und 21 Freiheitsgraden ist 3,47. Also ist  $H_0$  zu verwerfen und  $\beta_1$  und  $\beta_2$  sind vermutlich nicht beide Null (was nicht überrascht weil ja auch jeder Regressor einzeln, also für sich genommen, nach t-Test signifikant ist). Der *Gesamtzusammenhang* ist gesichert.<sup>8</sup>

##### b) Einfache Regression als Spezialfall

Wieder Affenaufgabe als Beispiel. Das Ergebnis des F-Tests ist natürlich das gleiche wie beim t-Test.  $S_{\hat{y}\hat{y}}$  ist zu errechnen als  $r^2 \cdot S_{yy} = 0,7857 \cdot 1628,8 = 1279,8$ .

<sup>7</sup> Terminologie: eine Variation ist eine Summe von Abweichungsquadraten (Abweichungen einer Größe von ihrem Mittelwert). Eine Varianz ist eine Variation dividiert durch die Anzahl der Freiheitsgrade. Es gibt bei der obigen Art der Varianzanalyse nur zwei Variationsquellen: erklärt (durch die Regressionsgleichung) und nicht erklärt (Störgröße).

<sup>8</sup> So, wie man bei einem Regressor ein Konfidenzintervall mit der t-Verteilung bestimmt, so kann man auch für zwei Regressoren gemeinsam eine Konfidenzellipse mit der F-Verteilung bestimmen (oder allgemein ein Konfidenzellipsoid bei  $K \geq 2$  Regressoren). Auch hier ist der Fall eines Regressors (Konfidenzintervall, t-Verteilung) wieder ein Spezialfall des Falles mehrerer Regressoren (Konfidenzellipsoid, F-Verteilung).

Variation	Freiheitsgrade	MSE
<b>erklärt</b> $S_{\hat{y}\hat{y}} = 1279,8$ (Regression)	$K = 1$	1279,8
<b>residual</b> $S_{\hat{u}\hat{u}} = 349$ (nicht erklärt)	$T-K-1 = 3$	$116,33 = \hat{\sigma}_u^2$

**Beachte:** K ist hier die Anzahl der Regressoren, nicht die Anzahl der zu schätzenden Parameter. Die ist  $k = K+1$ . Mit k wären die Freiheitsgrade entsprechend  $k-1$  und  $T-k$ .

$F = 11,001146$  wie man sieht, ist das gleich  $t^2$  denn  $t = 3,3173$ . Wenn es nur einen Regressor x gibt, ist der "Gesamt"-Zusammenhang natürlich dann gesichert, wenn der eine Regressor signifikant ist, also beim t-Test  $H_0$  abgelehnt wird.

### 5. Der F-Test für eine Teilgruppe von Regressoren (Signifikanz hinzukommender Regressoren) (= Fall 2b)

Es ist zu unterscheiden: das unrestringierte Modell (mit allen Regressoren), etwa  $\hat{y} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_5 x_5$  ( $K = 5$ ) und das restringierte Modell (bei Geltung von  $H_0$ ), wenn alle hinzukommenden Regressoren (zusammen) *nicht* signifikant sind, etwa (als Beispiel)  $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$  (es gilt  $L = 4$  Restriktionen und  $R\beta = q$ )

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ die Matrix } \mathbf{R} \text{ ist hier gleich der Einheitsmatrix } \mathbf{I}.$$

Die Summe der Quadrate der Abweichungen (Residuen) ist größer beim restringierten Modell als beim vollen (nicht-restringierten) Modell, denn wenn die hinzugenommenen Regressoren eine Erklärungskraft haben, wirkt sich das dahingehend aus, dass die Summe der quadrierten geschätzten Werte der Störgröße  $\hat{u}$  verringert. Die F-verteilte Prüfgröße lautet jetzt

$$F = \frac{(S_{\hat{u}\hat{u}}^0 - S_{\hat{u}\hat{u}}) / L}{S_{\hat{u}\hat{u}} / (T - K - 1)} \sim F_{L, T-K-1} \text{ und es gilt } S_{\hat{u}\hat{u}}^0 - S_{\hat{u}\hat{u}} = S_{\hat{y}\hat{y}} - S_{\hat{y}\hat{y}}^0$$

$S_{\hat{u}\hat{u}}^0$  ist die Summe der quadrierten Abweichungen im *restringierten* Modell (bei Geltung von  $H_0$ ) und  $S_{\hat{u}\hat{u}} = \sum \hat{u}^2$  ist die entsprechende Summe im *vollen* Modell (mit *allen* Regressoren).

$S_{\hat{u}\hat{u}}^0 - S_{\hat{u}\hat{u}} = S_{\hat{y}\hat{y}} - S_{\hat{y}\hat{y}}^0$  gilt weil  $S_{\hat{y}\hat{y}} + S_{\hat{u}\hat{u}} = S_{yy}$  beim vollen Modell und in gleicher Weise auch  $S_{\hat{y}\hat{y}}^0 + S_{\hat{u}\hat{u}}^0 = S_{yy}$  beim restringierten Modell.<sup>9</sup> Die F-Statistik hat deshalb die zusätzliche (durch hinzukommende Regressoren) erklärte Variation im Zähler.

Das folgende Zahlenbeispiel ist wieder dem Buch von Murphy entnommen ( $T = 50$ ):

<sup>9</sup> Die (nicht erklärte) Residualvariation nimmt somit im gleichen Maße ab wie die erklärte Variation zunimmt, wenn man vom restringierten Modell (ohne die Regressoren ...) zum vollen Modell übergeht. Die Differenz wäre nur dann Null, wenn wie in  $H_0$  angenommen wird, die zusätzlichen Regressoren keinen Erklärungsbeitrag liefern würden.

Variation	df*	MSE
<b>erklärt</b> $S_{\hat{y}\hat{y}} = 198$ (volles Modell, alle Regressoren)	$K = 5$	$198/5 = 39,6$
<b>residual</b> $S_{\hat{u}\hat{u}} = 134,7$ (nicht erklärt im vollen Modell)	$T - K - 1 = 44$	$134,7/44 = 3,06$
<b>erklärt</b> im restringierten** Modell $S_{\hat{y}\hat{y}}^0 = 158,2$	$K^0 = 1$	$158,2$
<b>residual</b> im restringierten Modell $S_{\hat{u}\hat{u}}^0 = 174,5$	$T - K^0 - 1 = 48$	$3,635$

\* df = degrees of freedom (Anzahl der Freiheitsgrade)

\*\* ohne die vier Regressoren  $x_2, \dots, x_5$ , also nur mit  $x_1$

Man verifiziert leicht, dass gilt  $S_{\hat{y}\hat{y}} = 198 + 134,7 = 158,2 + 174,5 = 332,7$  und man erhält auch  $S_{\hat{y}\hat{y}} - S_{\hat{y}\hat{y}}^0 = 198 - 158,2 = 39,8$  und  $S_{\hat{u}\hat{u}}^0 - S_{\hat{u}\hat{u}} = 174,5 - 134,7 = 39,8$  und  $L = K - K^0 = 4$ . Die Prüfgröße ist dann  $F = \frac{39,8/4}{134,7/44} = \frac{9,95}{3,06} = 3,25 \sim F_{4,44}$ .

F-Tabelle bei 95% und 99% Sicherheit und 4 und 44 Freiheitsgraden

95%:	2,58	99%:	3,78
------	------	------	------

Die zu  $x_1$  (restringiertes Modell gem.  $H_0$ ) hinzugekommenen Regressoren  $x_2, \dots, x_5$  liefern einen signifikanten Erklärungsbeitrag ( $H_0$  wird verworfen) auf dem 5%, nicht aber auf dem 1% Signifikanzniveau, denn die errechnete Prüfgröße  $F$  ist größer als der Tabellenwert bei 5% ( $3,25 > 2,58$ ).  $H_0$  ist aber nicht zu verwerfen auf dem 1% Niveau, denn bei 1% ist der Tabellenwert der F-Verteilung  $3,78 > 3,25$ .

### Zwei Ergänzungen:

F-Test im restringierten Modell ( $T - K - 1 = 50 - 1 - 1 = 48$ )

Der Regressor  $x_1$  ist im restringierten Modell signifikant, denn dann ist  $F_0 = \frac{158,2}{3,635} = 43,52$

$\sim F_{1,48}$ . Das Ergebnis ist gleich dem, was man mit dem t-Test erhält mit  $t = \sqrt{43,52} = 6,597$ . Die Tabellenwerte für die Signifikanzniveaus 5% und 1% sind

	5 %	1 %
F-Verteilung ( $F_{1,48}$ )	4,04	7,19
t-Verteilung ( $t_{48}$ )	1,67	2,39

Der (einzige) Regressor  $x_1$  ist also hochsignifikant (signifikant auf dem 1% Niveau)

Alle fünf Regressoren im vollen Modell mit  $x_1, x_2, \dots, x_5$  sind zusammen genommen nicht alle irrelevant, d.h. die Nullhypothese  $H_0: x_1 = x_2 = \dots = x_5 = 0$  wird verworfen, weil gilt

$$F_1 = \frac{198/5}{134,7/44} = \frac{39,6}{3,06} = 12,935 \sim F_{5,44}$$

Die Tabellenwerte der F-Verteilung sind 2,43 (5%) bzw. 3,46 (1%).

Man sieht, dass die Betrachtung dieses Abschnitts in die Abschn. 4 übergeht, wenn  $L = K$ , also die  $H_0$  lautet **alle** Regressoren sind zusammengenommen nicht signifikant.